Spring 2014 Genomics Exam #2
Sequence Variations and Microarrays

There is no time limit on this test, though I don't want you to spend too much time on it. I have tried to design an exam that will take less time than exams in the past. You do not need to read any additional papers other than the ones I send to you. There are 5 pages, including this cover sheet, for this test. You are not allowed discuss the test with anyone until all exams are turned in no later than 10:30 am on Wednesday March 26. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 10:30 am ON WEDNESDAY MARCH 26**. You may use a calculator, a ruler, your notes, the book, and the internet. You may take this exam in as many blocks of time as you want. Submit your electronic version before 10:30 am (eastern time zone).

The **answers to the questions must be typed in a Word file and emailed to me as an attachment**. Be sure to backup your test answers just in case (I suggest a thumb drive or other removable device). You will need to capture screen images as a part of your answers which you may do without seeking permission since your test answers will not be in the public domain. Remember to explain your thoughts in your own words and use screen shots to support your answers. Screen shots without ***your*** words are worth very few points. Support your answers with data using screen shots liberally (no permission required since your exam is a private document).

DO NOT READ or DOWNLOAD ANY PAPERS FOR THIS EXAM. RELY ON YOUR EXPERIENCE, AND YOUR SKILLS.

-3 pts if you do not follow this direction.
Please do not write or type your name on any page other than this cover page.
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please type):

Write out the full pledge and sign (electronic signature is ideal):

How long did this exam take you to complete?

40 points

1) By now, you know how much I enjoy learning and sharing with my students. In preparing for this exam, I stumbled on a biological reality that rocked my perception of the genome and cells (Figures 1 and 2).

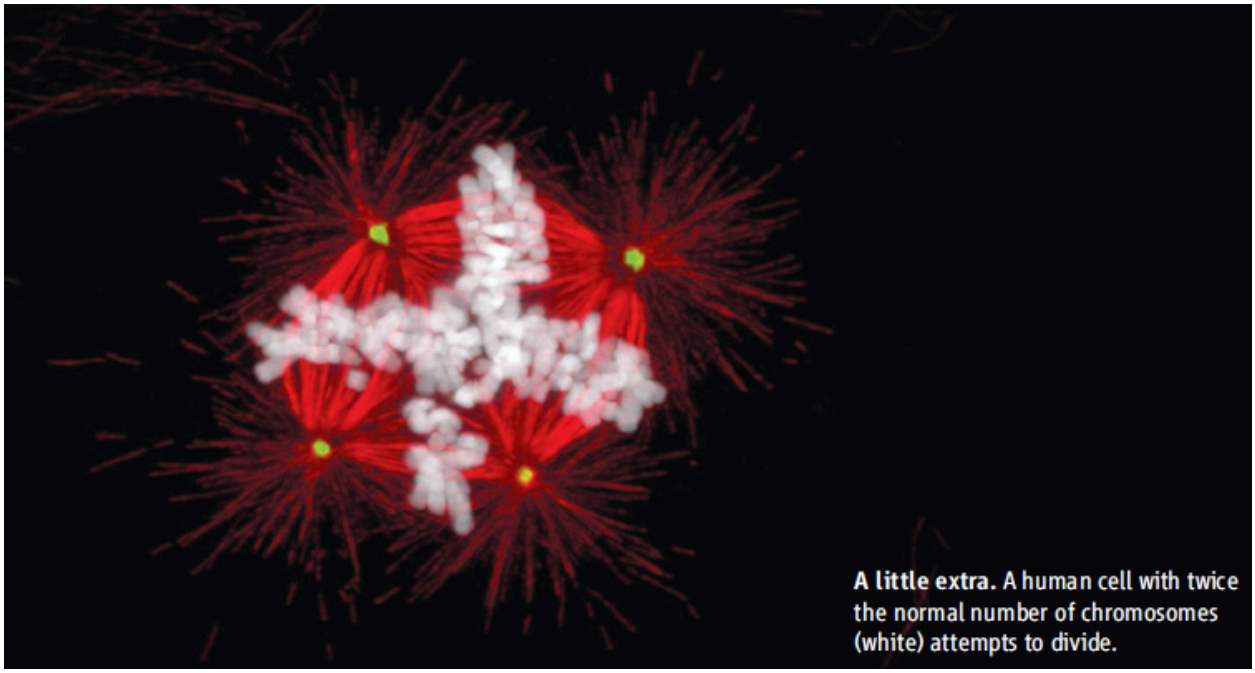
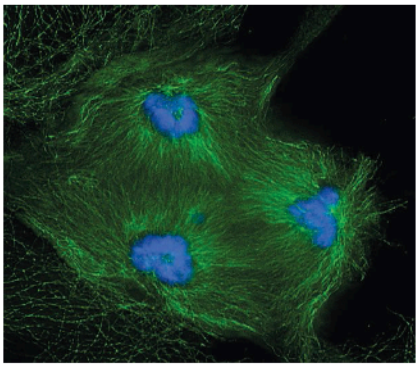


Figure 1. Image that blew Dr. C's world view.

POLYPLOID CELL TYPES IN MAMMALS

CELL TYPE	LOCATION	FUNCTION	NUMBER OF GENOME COPIES
Megakaryocyte	Bone marrow	Producing blood clotting platelets	Up to 128
Hepatocyte	Liver	Detoxification, metabolism	Typically 4 to 16
Trophoblast giant cell	Embryo	Promote implantation	Up to 1000
Cardiomyocyte	Heart	Contraction	Typically 4



Bonus DNA. The polyploid cells in mammalian bodies differ in their location, function, and number of chromosome sets (table). In a liver cell (*right*), the multiple chromosome copies (blue) have sorted into three clusters in preparation for cell division.

Figure 2. Information that made Dr. C. realize he had been lied to for decades about humans as diploids organisms.

5 points

a) Find a human gene that has experimental evidence supporting its role in normal (non-cancerous) polyploidy formation in human cells. You must provide two independent sources that verify your gene is involved.

AKT1: V-akt murine thymoma viral oncogene homolog 1

5 points

b) determine the expression of your gene in a wide range of human tissues. Show your data and list the top 3 expressing healthy tissues.
http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Hs.525622

10 points

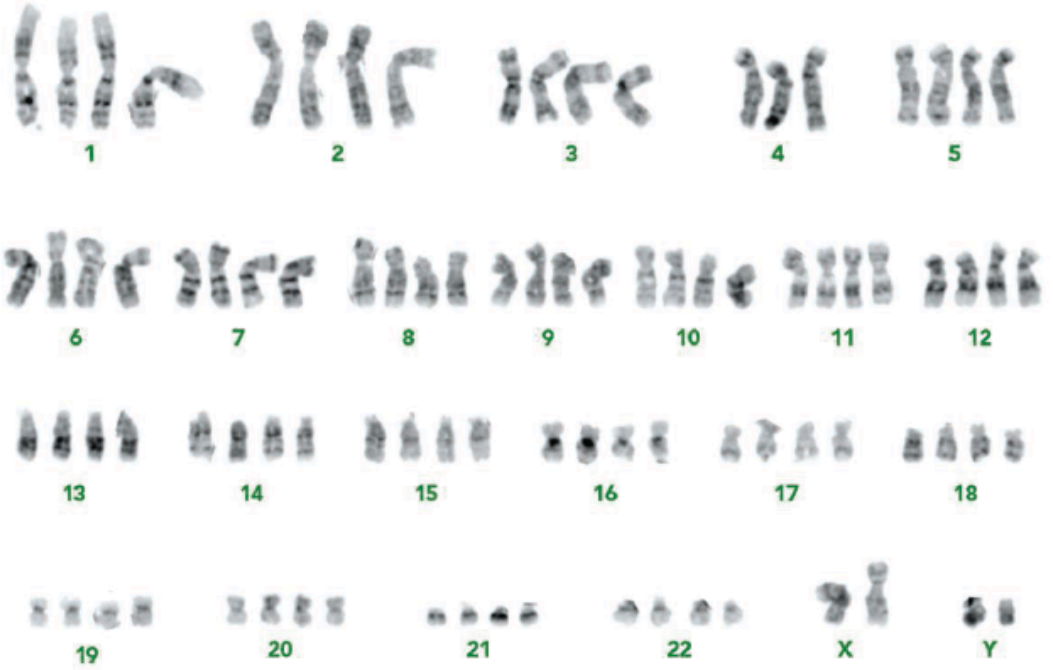
c) For the gene you found in part (a) above, find evidence of two categories of sequence variations in humans. One source must show mutations that lead to disease states. The second source must show variations that provide population frequency information for a range of ethnically diverse people but these mutations cause no known negative phenotypes.

http://www.ncbi.nlm.nih.gov/clinvar?LinkName=gene_clinvar&from_uid=207

http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/?chr=NC_000014.9&from=104769348&to=104795742

20 points

d) Look at this karyotype from a non-pathogenic human liver cell.



Double down. The chromosome copies from a polyploid liver cell arranged by size, showing that the cell carries four copies of almost every one.

Design an experiment to determine what percentage of liver cells are tetraploid. Your experimental method should also be able to detect when one or more chromosomes are not fully

tetraploid as you can see for chromosome 4 above. **Using an outline format**, tell me what you would do starting with a 5 gram piece of liver. What methods would you use and what sorts of data would you collect? Give me some theoretical data that you would expect to get from your method.

Two high-throughput methods:
tissue slices with FISH probes, one for each chromosome OR DNA sequencing of single cells with known ploidy such as sperm cells which are known haploid and easier to obtain than eggs

transcriptome was less than ideal on microarrays since you are making assumptions about transcription being proportional to chromosome number. A better microarray method would have been comparative genome hybridization but still limited by microarray underestimation of quantity.

35 points

2) There are many aspects of biology that fascinate me. One is the creativity of investigators who either ask questions that never occurred to me. Here is some text from a recent abstract that poses a question that would never occur to me.

> Humans traveling in space might allow us unprecedented exploration and discovery, but we need to more fully understand the consequences of long-term exposure to spaceflight. Microgravity (μg) is constant in space, and hypergravity (hyper g) is experienced during launch and landing. Immune dysfunction in both μg and hyper g has been independently documented multiple times. The human immune system is weakened in prolonged exposure to space travel which results in increased vulnerability to opportunistic infections. To understand the human immune system when exposed to space exploration, we need to better understand how the metazoan immune systems response to space flight. We used Drosophila as a model of innate immunity to see how space flight affects a normal immune response to pathogens as simulated by exposure to the common immunogen

5 points

a) Summarize the results from Figure 2a. Support the main conclusion using data in this figure.
A gene called *yuri* is responsible for sensing gravity in flies. Mutnant fly strain *yuri* has a deleted *yuri* gene while the fly UAS contains the same deleted *yuri* gene as well as a transgene of *yuri* that replaces the missing gene and encoded protein. 1 g = earth gravity; 4 g = hyper g; minus sign indicates non-infected flies; + indicates flies infected by a fungus. Error bars = SEM for 3 experiments.

top: fungal infection in yuri- strain of flies was no different in regular and hyper g.
middle: over expressing yuri reduced fly mortality from fungal infection at hyper g. It appears immune function is improved in hyper g compared to normal g.
bottom: confirms that hyper g is protective for fly defense from fungal infection. Also confirms that yuri is a key gene in the connection between hyper g and improved immunity.

5 points

b) From Figure 2a, can you distinguish between the two possibilities that either 1) the fungus has an altered virulence at hyper g or 2) the fly has an altered immune response to hyper g? Explain how you reached your conclusion.

as noted above, wt and UAS both had improved immunity in hyper g but in the yuri- fly, the virulence was the same in 1g and 4g. Therefore, the fungus does not have higher virulence at 4 g since it was the same as in 1g in the absence of yuri.

5 points

c) List the signature gene sets that are induced or repressed only in one gravity environment with a P value of at least 10^{-6} in Figure 2b.

Fly DNA microarrays were used to survey every gene's activity comparing flies on earth vs flies

on the space station. The number of genes induced or repressed in 1g flies only (Earth) or μg flies only (Space) or in both (overlap) are listed in the Venn diagrams. Signature set genes are shown on the right side of the figure. Figures show the number of genes in each signature (bar graphs; upper axis), and the P values showing statistical over-representation of each signature set (circles; lower axis).

induced 1g gene sets
1. innate immune response
2. serine-type peptidase activity
3. response to fungus
4. toll signaling pathway
No induced or repressed gene expression in any of the other 3 categories with p value of at least 10^{-6}.

5 points

d) Figure 2c shows microarray data for 7 genes involved in innate immunity. Affymetrix single channel microarrays were used to generate the data. * = $p < 0.01$; ** = $p < 0.005$; *** = $p < 0.001$. Produce a Table that has the name of each gene at the top of columns and the rows are labled by the experimental conditions. Then fill in the table by typing "ind" for every gene that is induced or "rep" for every gene that is repressed on earth compared to in space. U = uninfected flies; B = bacterially infected flies; F = fungally infected flies. Leave table cells blank if they are not altered significantly.

	pelle	cactus	Metchnikowin	Drosomycin	necrotic	toll	Drosomycin-like5
Un1g							
Bac1g			ind	ind	rep		
Fun1g	ind	ind	ind	ind	ind	ind	ind
Un μg							
Bac μg			ind	ind	ind		
Fun μg							ind

Red text shows gene regulation on 1 g (Earth) compared to μg (space).
Black text shows gene regluation within one gravity compared to unifected.

5 points

e) Look at Figure 2d. As you did for question 2c above, which signature gene set show significan differences? U = uninfected; E = earth; S = space station; 1 – 3 represent the triplicate batches of flies. Color scale uses the same trends that Botstein and Brown developed for their microarrays.
If we apply the same P value (10^{-6}), then the only gene set induced is stress response genes in μg.

5 points

f) All of the stress response genes in Figugre 2d are heat shock protein (hsp) paralogs. Look up what hsp do with regards to protein folding. Speculate why hsp gene regulation is an idication of what is going wrong with fly innate immunity with regards to gravity. Support your speculation with data presented in this exam as well as onlie resources.
In microgravity, it looks like protein folding may not be functioning normally, so the hsp may be induced to help proteins take the right conformation. This would also suggest that fungal proteins may fold better in microgravity than fly proteins involved in innate immunity.

5 points
g) What was the value of the DNA microarray study?
It pointed out that the innate immune response to fungus, via toll receptor maybe, is activated at 1g when infected with a virus. This hints at the improved immune response in hyper g might be due to the same genes as listed in part c above. It was way to find immune response genes to test further for response to gravity.

25 points

3)

5 points

a) Summarize the genomic and evolutionary history of the bonobo and chimpanzee. Your answer cannot exceed 75 words. Cite your source of information (citation is not part of the word count). You may want to refer to Figure 3map for an understanding of geography.

http://www.eva.mpg.de/3chimps/files/apes.htm

Bonobos and chimpanzees diverged from each other around 2 million years ago and differ in morphology, behavior, and perhaps even emotions and cognition in important ways.

Bonobos and Chimpanzees share close to 99% of their genome in common with humans, meaning that their genomes are more similar to that of humans than they are to that of gorillas. However, it may be that Bonobos, whose psychology is virtually unstudied relative to that of chimpanzees, are more similar to humans than are chimpanzees in how they solve various social problems (e.g. Hare, Melis, Woods, Hastings, & Wrangham, 2007). Such similarities may even be partly the result of shared and heritable neurophysiology that potentially regulates the social emotions of humans and Bonobos in similar ways (Hammock & Young, 2005).

http://en.wikipedia.org/wiki/Bonobo

DNA evidence suggests the bonobo and common chimpanzee species effectively separated from each other fewer than one million years ago.[22][23] The *Pan* line split from the last common ancestor shared with humans approximately six to seven million years ago. Because no species other than *Homo sapiens* has survived from the human line of that branching, both *Pan* species are the closest living relatives of humans and cladistically are equally close to humans. The recent genome data confirms the genetic equidistance.

5 points

b) The investigators sequenced the mtDNA as well as 15 different autosomal loci. Figure 3a shows you the evolutionary tree from a subset of these data. Describe the overall genome variation for the five different populations whose DNA were sequenced. Red = bonobos; green for western chimps; grey = central; blue = eastern; yellow = Nigerian/Cameroonian (aka vellerosus).

- **bonobos** (red) are tightly clustered therefore similar to each other. They are distinct from the others in genes b, c and mitochondrial, but not so for nuclear gene a (grey and blue are with red).
- **western** (green) chips are also tightly clustered and vellerosus (yellow) seem most similar to western.
- **central, eastern and vellerosus** (gene c only) are less homogeneous in these 3 nuclear genes and one mitochondrial gene.

5 points

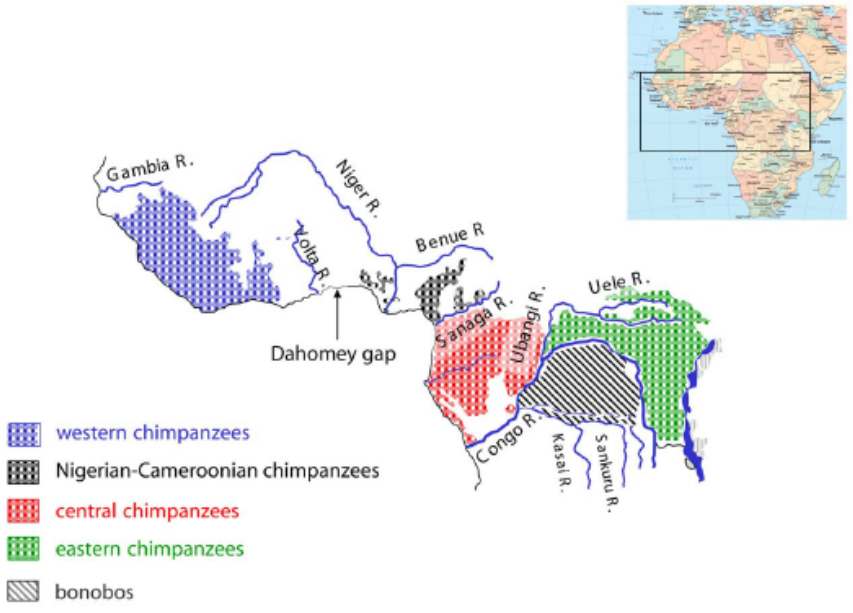
c) We used the F_{st} statistic once in class. Apply what you know from that example to Figure 3b. Summarize your interpretation of this comparative table.

This comparison of genetic variability within and between populations is frequently used in applied population genetics. The values range from 0 to 1. A 0 value implies that the two populations are interbreeding freely. A value of 1 implies that the two populations do not share any genetic diversity. (http://en.wikipedia.org/wiki/Fixation_index)

Western chimp and bonobos are the most isolated (0.74) while central and eastern are the most closely related due to interbreeding (0.07).

Nigerian and central also interbreed a lot (0.16) which makes sense given their proximity.

You might think bonobo and eastern would interbreed a lot based on geography but they are no more related (0.56) than the Nigerian chimps which are geographically separated. The Congo River separates eastern from bonobo.



5 points

d) Look at Figure 3c which uses principle component analysis. What can you conclude about this figure? Is it consistent or inconsistent with what you have learned so far in Question 3? Here are the colors for Figure 3c: green for bonobos; orange, western; blue, central; red, eastern; black Nigerian-Cameroonian.

Yes, this is very consistent with the Fst values in the previous figure. Western high on both axes and bonobo low on PC1 medium on PC2. All others fairly tight cluster.

5 points

e) Change of pace: Some folks at Carnegie Mellon University developed a meta-analysis tool (ExpressionBlast) for transcriptome data. Choose their sample project to find a human gene that is down regulated when exposed to 25 μM resveratrol and show me the screen shot of that gene.

http://www.expression.cs.cmu.edu/index.html